

EDF 기대손실에 기초한 공정능력지수의 붓스트랩 신뢰구간

임태진 · 송현석

송실대학교 산업·정보시스템공학과

Bootstrap Confidence Intervals of the Process Capability Index Based on the EDF Expected Loss

Lim, Tae-Jin · Song, Hyun-Suk

Dept. of Industrial Information System Engineering, Soongsil University

Key Words : Bootstrap, EDF, PCI, confidence interval, reflected normal loss function

Abstract

This paper investigates bootstrap confidence intervals of the process capability index(PCI) based on the expected loss derived from the empirical distribution function(EDF). The PCI based on the expected loss is too complex to derive its confidence interval analytically, so the bootstrap method is a good alternative. We propose three types of the bootstrap confidence interval; the standard bootstrap(SB), the percentile bootstrap(PB), and the acceleration biased-corrected percentile bootstrap(ABC). We also perform a comprehensive simulation study under various process distributions, in order to compare the accuracy of the coverage probability of the bootstrap confidence intervals. In most cases, the coverage probabilities of the bootstrap confidence intervals from the EDF PCI turned out to be more accurate than those from the PCI based on the normal distribution. It is expected that the bootstrap confidence intervals from the EDF PCI can be utilized in real processes where the true distribution family may not be known.

1. 서론

품질의 척도로써 공정능력지수(Process Capability Index: PCI)의 사용이 증대되고 있으나 전통적인 공정능력지수는 단지 품질

의 산포만을 고려한 지수이기 때문에 고객의 입장에서 경제적 손실을 고려한 평가를 수행하기 어려웠다. 이러한 요구에 의해 Taguchi의 이차 손실함수를 이용한 공정능력지수 C_{pm}^+ 가 개발되어 비용적 측면까지

고려한 공정평가가 가능하게 되었다 [Boyles, 1991] [Chan et al., 1988]. 그러나 Taguchi의 이차 손실함수 또한 규격한계에 관계없이 이차 함수를 적용한다는 단점이 있어 이를 보강하기 위한 역정규 손실함수 (Inverted Normal Loss Function : INLF)가 개발되어 현실적으로 규격한계에 따른 최대 손실 적용이 가능하게 되었다 [Spiring, 1993]. 역정규 손실함수의 개발 이후로 역정규 손실함수를 사용한 기대손실에 대한 연구가 진행되었다 [Spiring & Yeung, 1998]. 정영배와 문혜진(2002)은 Spiring(1993)의 역정규 손실함수에 기초한 공정능력지수 C_{pl} 를 제안하고, 다양한 모의실험을 통해 기존 공정능력지수와의 비교·연구를 수행하였다. 그러나 지수 C_{pl} 는 공정분포가 정규분포가 아닌 경우에는 함수식의 복잡성으로 인하여 정확한 기대손실을 도출해내기가 어렵고, 실제 산업 현장에서 나타낼 수 있는 비대칭 손실함수에 대한 정확한 기대손실을 계산하기 어렵다. 또한 정규분포를 따르지 않는 공정에 대해서는 그 특성이 밝혀지지 않았다. 이에 착안하여 임태진과 송현석(2003)은 실험적 분포(Empirical Distribution Function: EDF)를 이용한 공정능력지수 C_{pm} 을 개발하여, 비대칭 손실함수에 대해서도 계산이 용이하며, 다양한 공정분포에 대하여 강건하고(robust), 정확함을 모의실험을 통해 입증하였다.

새로운 공정능력지수의 개발 못지않게 중요한 것이 추정치의 참값에 대한 신뢰구간을 추정하는 문제이다. 전통적인 공정능력지수의 신뢰구간에 대한 해석적인 연구가 있었으나 [Chan et al., 1990] [Pearn et al., 1992], 최근 개발된 지수 C_{pl} 나 C_{pm} 은 수식의 복잡성으로 인하여 해석적인 접근이

어려운 실정이다.

붓스트랩 방법은 C_{pl} 나 C_{pm} 와 같은 복잡한 통계량에도 적용이 가능하고, 정규성의 가정 없이도 비교적 정확하게 신뢰구간들을 추정해 주는 것으로 알려져 있다 [조중재 등, 1999] [Choi et al., 1996] [Franklin & Wasserman, 1992].

본 논문에서는 공정능력지수 C_{pl} 및 C_{pm} 의 신뢰구간을 표준 붓스트랩(Standard Bootstrap: SB), 백분위수 붓스트랩(Percentile Bootstrap: PB), 및 가속 편이수정 백분위수 붓스트랩(Acceleration Biased-Corrected Percentile Bootstrap: ABC) 등 세 가지 방법을 통하여 추정하고 그 정확성을 비교한다. 정규분포, 카이제곱분포, 스튜던트- t 분포, 일양분포 등의 공정분포에 대하여 각각 붓스트랩 신뢰구간을 추정하고 포함(coverage) 확률을 계산하여 신뢰구간의 정확성을 평가함으로써, EDF 지수 C_{pm} 의 강건성(robustness)과 붓스트랩 신뢰구간의 정확성을 보이고자 한다.

2장에서는 기대손실을 이용한 공정능력지수 및 EDF에 기초한 지수 C_{pm} 의 개념을 설명한다. 3장에서는 세 가지 붓스트랩 신뢰구간에 대해서 설명하고, 4장에서는 모의실험을 통하여 지수 C_{pl} 및 C_{pm} 에 대한 각 신뢰구간의 정확성을 비교한다.

2. 기대손실을 이용한 공정능력지수

2.1 이차 기대손실함수를 이용한 공정능력지수

전통적인 공정능력지수들은 공정의 산포 및 중심위치를 근거로 하여 공정능력을 평가한다. 기대손실함수를 이용하여 공정능력을 평가하는 것은 목표치로부터의 품질 변동에 따른 경제적 손실까지 고려한다는 점에서 그 의의를 찾을 수 있다. Taguchi의 이차손실함수는 품질 특성치가 목표치와 일치할 때는 손실이 발생하지 않으며, 점차 멀어질수록 손실이 크게 발생한다는 가정 하에서 이차식으로 근사화한 손실함수이다. 이차손실함수를 $L(x, T)_t$ 라 하고 그 기대치를 $E[L(x, T)]_t$ 라 하면, 품질의 특성치가 망목 특성일 때 각각 식 (1)과 (2)와 같이 나타낼 수 있다.

$$L(x, T)_t = k(x - T)^2 \tag{1}$$

$$E[L(x, T)]_t = k[\sigma^2 + (\mu - T)^2] \tag{2}$$

단, $k = \frac{A}{\Delta}$ 이며 A 는 품질 특성치가 가질 수 있는 최대 손실이고, Δ 는 최대 손실 A 가 첫 번째로 발생하는 시점에서의 목표치로부터 떨어진 정도(거리)를 나타낸다.

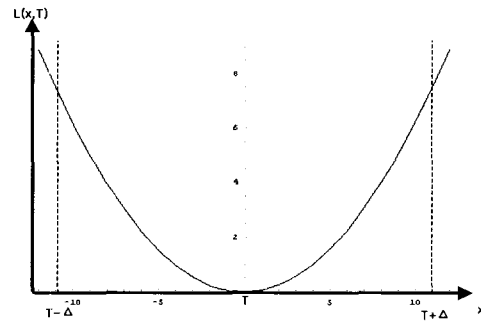
Chan et al.(1988)과 Boyles(1991)는 각각 독립적인 연구를 통하여 Taguchi 이차손실함수의 기대손실을 이용한 공정능력지수 C_{pm}^+ 를 다음과 같이 제안하였다.

$$C_{pm}^+ = \frac{USL - LSL}{6\sqrt{E[L(x, T)]_t}} \tag{3}$$

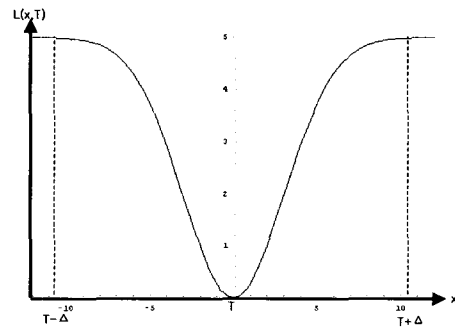
2.2 역정규 기대손실함수를 이용한 공정능력지수

Taguchi의 이차손실함수는 공정이 정규 분포인 경우에만 잘 설명이 되고, <그림

1>에서 보는 바와 같이 품질의 특성치가 규격을 벗어났을 때에도 손실이 계속 증가하는 것으로 해석되는 단점이 있다.



<그림 1> Taguchi의 이차 손실함수



<그림 2> 역정규 손실함수

이런 단점을 보완하여 Spiring(1993)은 <그림 2>와 같은 역정규 손실함수(INLF)를 개발하였다. INLF는 기존의 이차 손실함수에 비해 목표치로부터 품질의 변동에 따른 손실을 보다 합리적으로 설명할 수 있고 응용의 폭도 넓힐 수 있는 장점이 있다. 또한 <그림 2>에서 보는 바와 같이 INLF는 '항상 양수의 값을 갖고, 목표치에서 가장 최소의 값을 가지며, 목표치에서 멀어질수록 공정의 분산은 커지고 그에 따른 손실도 증가한다'는 손실(loss)의 일반적인 특성에 잘 부합되면서, Taguchi의 이차손실함수

와는 달리 규격을 벗어날 때 최대손실 이상의 값을 가지지 않는다.

Spiring & Yeung(1998)은 INLF를 일반화하여 다음과 같이 비정규분포에도 적용 가능한 역분포 손실함수를 개발하였다.

$$L(x, T) = A \left(1 - \frac{g(x, T)}{m} \right) \quad \forall x \in \Omega \quad (4)$$

단, $g(x, T)$ 는 손실함수에 이용되는 p.d.f., m 은 함수 $g(x, T)$ 가 최대도 도달할 수 있는 값, A 는 손실함수가 갖는 최대의 손실, Ω 는 x 값이 가지는 범위로써 $(-\infty, \infty)$ 등이다.

공정 품질측정치 분포의 p.d.f.를 $f(x)$ 라 하면, 기대손실은 다음과 같이 계산된다.

$$E[L(x, T)] = \int_{\Omega} A \left(1 - \frac{g(x, T)}{m} \right) dF(x) \quad (5)$$

대칭 규격이 주어진 경우, 가장 일반적으로 사용되는 역분포로는 다음과 같이 목표치 T 를 중심으로 하는 정규분포이다.

$$g(x, T) = \frac{1}{\sqrt{2\pi}\gamma} \text{Exp} \left\{ -\frac{(x-T)^2}{2\gamma^2} \right\} \quad (6)$$

단, T 는 목표치, γ 는 형상모수, $\gamma = \frac{\Delta}{4}$,

Δ 는 최대 손실 A 가 첫 번째로 발생하는 시점에서의 목표치로부터 떨어진 정도(거리)가 된다.

품질의 특성치 x 가 목표치 T 와 일치할 때 $g(x, T)$ 가 최대가 되므로 $m = \frac{1}{\sqrt{2\pi}\gamma}$

이고, 이를 식 (6)에 대입하면 손실함수는 식 (7)과 같이 정의된다.

$$\begin{aligned} L(x, T) &= A \left[1 - \frac{1}{\sqrt{2\pi}\gamma} \text{Exp} \left\{ \frac{-(x-T)^2}{2\gamma^2} \right\} / \left(\frac{1}{\sqrt{2\pi}\gamma} \right) \right] \\ &= A \left(1 - \text{Exp} \left\{ \frac{-(x-T)^2}{2\gamma^2} \right\} \right) \quad (7) \end{aligned}$$

공정의 분포가 정규분포일 때의 역정규 손실함수에 기초한 기대손실은 식 (5)와 (7)로부터 다음과 같이 계산된다[Spiring, 1993].

$$E[L(x, T)]_I = A \left[1 - \frac{\gamma}{\sqrt{\sigma^2 + \gamma^2}} \text{Exp} \left\{ \frac{-1}{2} \left(\frac{\mu - T}{\sigma^2 + \gamma^2} \right)^2 \right\} \right] \quad (8)$$

정영배와 문혜진(2002)은 역정규 손실함수의 기대손실 $E[L(x, T)]_I$ 에 기초한 공정능력지수 C_{pl} 를 다음과 같이 제안하였다.

$$C_{pl} = \frac{USL - LSL}{6\sqrt{E[L(x, T)]_I}} \quad (9)$$

지수 C_{pl} 는 기존의 이차손실함수에 기초한 지수 C_{pm}^+ 보다 공정 표준편차의 변화에 대해서 민감하게 반응하는 것으로 나타났다[정영배와 문혜진, 2002]. 그러나 공정이 정규분포를 따르지 않는 경우에는 기대손실 자체의 오류로 인하여 정확성에 문제가 발생할 수 있다.

2.3 EDF를 이용한 기대손실과 공정능력지수

분포함수 $F(x)$ 로부터 샘플링한 확률표본 X_1, X_2, \dots, X_n 의 관측값을 x_1, x_2, \dots, x_n 이라 하면, 선형적 분포함수(EDF)는 다음과 같이 정의된다.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \quad (10)$$

EDF는 특정 분포를 가정하지 않고, 단순히 주어진 값 x 를 넘지 않는 관측값의 비율로서 계산되므로, 분포에 대한 강건성을 기대할 수 있다.

공정의 분포에 대한 가정을 배제하고, 식

(10)의 EDF를 식(5)에 대입하면 다음과 같이 기대손실을 계산할 수 있다.

$$\begin{aligned} E[L(x, T)] &= A \left[1 - \frac{1}{m} \int_{-\infty}^{\infty} g(x, T) dF_n(x) \right] \\ &= A \left(1 - \frac{1}{m} \sum_{i=1}^n g(x_i, T) \frac{1}{n} \right) \quad (11) \end{aligned}$$

망목특성에 적합한 역정규 손실함수는 식(6)과 같고, 이를 식(11)에 대입하면 다음과 같이 기대손실을 얻는다.

$$\begin{aligned} E[L(x, T)] &= \int_{-\infty}^{\infty} A \left[1 - \frac{1}{\sqrt{2\pi\gamma}} \text{Exp} \left\{ \frac{-(x-T)^2}{2\gamma^2} \right\} / \left(\frac{1}{\sqrt{2\pi\gamma}} \right) \right] dF_n(x) \\ &= A \left(1 - \sum_{i=1}^n \text{Exp} \left\{ \frac{-(x_i-T)^2}{2\gamma^2} \right\} \frac{1}{n} \right) \quad (12) \end{aligned}$$

이와 같이 EDF를 공정의 분포로 하고 역정규 손실함수를 사용한 기대손실과 공정능력지수를 각각 $E[L(x, T)]_n$ 와 C_{pn} 으로 정의한다.

$$C_{pn} = \frac{USL - LSL}{6\sqrt{E[L(x, T)]_n}} \quad (13)$$

임태진 등(2003)은 정규분포, 스튜던트 t 분포, 카이제곱분포 등 다양한 분포에 대하여 모의실험을 수행하여 지수 C_{pn} 의 강건성을 보였다. EDF 지수 C_{pn} 은 계산이 간편하고, 공정분포 유형에 관계없이 정확하다는 장점이 있으나, 신뢰구간 계산이 용이하지 않다.

3. 붓스트랩 신뢰구간

최근 들어 공정능력지수의 신뢰구간 연구에 도입되고 있는 붓스트랩 방법은 컴퓨터의 계산 속도가 향상됨에 따라 유용한 비모수 통계적 기법으로 자리 잡고 있다. 공정능력지수에 대한 추정량의 분포적 특성의 어려움에 제약을 받지 않게 됨으로써, 공정능력지수의 신뢰구간을 추정하는데 적합하다.

붓스트랩 방법은 컴퓨터 기반 비모수적 몬테칼로 시뮬레이션 방법으로서 Efron(1979)에 의해 소개되어졌다. Hall(1988)은 다양한 붓스트랩 신뢰구간들을 포괄적으로 연구하였고, Franklin & Wasserman(1992)은 공정능력지수 C_p , C_{pk} , C_{pm} 등에 대하여 세 가지 붓스트랩 신뢰구간(SB, PB, BCPB)을 비교·분석하였다. Choi et al.(1996)은 C_p , C_{pk} , C_{pm} , C_{pmk} 등의 극한분포를 이용하여 Percentile- t 방법을 이용한 붓스트랩 신뢰구간을 제안하였다. 그들은 일반적인 붓스트랩 추정방법인 SB, PB를 Percentile- t 방법과 비교하여 Percentile- t 신뢰구간 추정방법이 더 우수함을 보였다. 조중재(1999)는 C_{pk} 에 대한 보다 나은 구간추정을 위하여 SB, STUD, HYB, BACK, BC, ABC 등의 6가지 추정방법을 가지고 포괄적인 비교실험을 수행하였다.

공정능력지수의 참값을 추정할 때에는, 샘플링 변동에 기인하는 불확실성을 고려하여 신뢰구간을 추정하는 것이 무엇보다 중요하다. 따라서 본 논문에서는 C_{pl} 와 C_{pn} 에 대한 표준 붓스트랩 신뢰구간, 백분위 신뢰구간, 편의가 수정된 백분위 신뢰구간 등을 각각 추정하고, 모의실험에 의해 포함

확률을 추정하여 각 공정능력지수를 비교·평가한다.

3.1 붓스트랩 샘플링 방법

원래의 데이터들의 집합 $\{x_1, x_2, \dots, x_n\}$ 에서 n 번 랜덤하게 복원 추출하여 얻어진 $\{x_1^*, x_2^*, \dots, x_n^*\}$ 을 붓스트랩 샘플이라고 한다. 예를 들어 원 데이터가 $n=5$ 인 $\{x_1, x_2, x_3, x_4, x_5\}$ 라 할 때, 랜덤하게 복원 추출하여 $\{x_5, x_5, x_1, x_2, x_4\}$ 와 같은 재추출된 붓스트랩 데이터 중 하나를 얻을 수 있다. 따라서 원래의 데이터를 가지고 계산된 공정능력지수를 \mathcal{C}_p 라하고 재추출된 데이터를 이용하여 만든 공정능력 지수를 C_p^* 라고 한다면 B 로 대표되는 붓스트랩 샘플 횟수만큼 C_p^* 를 만들어낼 수 있는데, n^n 의 조합까지 가능하다. 그러나 Efron(1993)은 신뢰구간 추정량들을 계산하기 위해서는 최소한 1000개의 붓스트랩 표본이면 대개 충분하다고 논하였다.

3.2 붓스트랩 신뢰구간 설정방법

여러 가지 붓스트랩 신뢰구간 설정방법의 설명을 용이하기 위해 기호를 정리하면 아래와 같다.

- B : 붓스트랩 횟수
- C_p^+ : 공정능력지수의 참값
- \mathcal{C}_p : 원데이터로 계산된 공정능력지수값
- $C_{p,i}^*$: i 번째 붓스트랩 데이터로 계산된 공

정능력지수값

$C_{p(k)}^*$: $C_{p,i}^*$ 들을 크기 순으로 정렬하였을 때 k 번째 값

3.2.1 표준 붓스트랩(SB) 신뢰구간

먼저 B 개의 붓스트랩 추정치 $C_{p,i}^*$ 들의 평균 \mathcal{C}_p^* 와 표준편차 S_c^* 를 각각 아래와 같이 계산한다.

$$\mathcal{C}_p^* = \frac{1}{B} \sum_{i=1}^B C_{p,i}^*$$

$$S_c^* = \sqrt{\frac{1}{B-1} \sum_{i=1}^B [C_{p,i}^* - \mathcal{C}_p^*]^2}$$

따라서 각 공정능력지수에 대한 $100(1-\alpha)\%$ 표준 붓스트랩 신뢰구간은 식 (14)와 같이 추정한다.

$$[\mathcal{C}_p - z_{\alpha/2} S_c^*, \mathcal{C}_p + z_{\alpha/2} S_c^*] \quad (14)$$

단, $1-\alpha$ 는 신뢰확률이고 z_p 는 표준정규 분포의 $(1-p)$ 분위수를 나타낸다.

3.2.2 백분위 붓스트랩(PB) 신뢰구간

백분위 붓스트랩 신뢰구간은 붓스트랩 데이터로 생성된 $C_{p,i}^*$ 들을 순서대로 정렬시켜 식 (15)와 같이 $[B \times (\alpha/2)]$ 와 $[B \times (1 - \alpha/2)]$ 번째 값을 각각 신뢰하한 및 상한으로 둔다.

$$[C_{p([B \times (\alpha/2)])}^*, C_{p([B \times (1 - \alpha/2) 1])}^*] \quad (15)$$

3.2.3 편의가 수정된 백분위 신뢰구간

ABC 방법은 앞에서 다룬 백분위 방법보다 복잡하지만 개선된 결과를 제공할 수 있다. ABC 방법에 의한 신뢰구간은 백분위 구간과 마찬가지로 백분위 값을 이용하지만 가속계수 \hat{a} 와 편향정정(bias-correction)

추정치 \hat{z}_0 를 활용한다. ABC 신뢰구간은 식 (16)과 같다.

$$[C_{p(B \times \alpha_1)}^*, C_{p(B \times \alpha_2)}^*] \quad (16)$$

$$\text{단, } \alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 - z_{(a)}}{1 - \hat{a}(\hat{z}_0 + z_{(a)})}\right),$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{(1-a)}}{1 - \hat{a}(\hat{z}_0 + z_{(1-a)})}\right),$$

Φ 는 표준정규 확률변수의 누적분포이고 \hat{z}_0 은 아래와 같이 계산된다[Efron, 1993].

$$\hat{z}_0 = \Phi^{-1}(\sum_{i=1}^B I(C_{p,i}^* \leq \mathcal{C}_p) / B)$$

단, $I(\cdot)$ 은 지시함수를 나타낸다.

가속계수 \hat{a} 은 잭나이프(jackknife) 방법에 의해 아래와 같이 계산한다[Efron, 1993].

$$\hat{a} = \frac{\sum_{i=1}^n (\mathcal{C}_{p<\cdot>}^* - \mathcal{C}_{p<i>}^*)^3}{6 \left\{ \sum_{i=1}^n (\mathcal{C}_{p<\cdot>}^* - \mathcal{C}_{p<i>}^*)^2 \right\}^{3/2}}$$

단, $\mathcal{C}_{p<i>}^*$ 는 원데이터에서 i 번째 관측값 x_i 를 제외하고 추정된 값을 나타내며, $\mathcal{C}_{p<\cdot>}^*$ 는 $\mathcal{C}_{p<1>}^*, \dots, \mathcal{C}_{p<n>}^*$ 들의 평균을 나타낸다.

편의가 수정된 백분위 신뢰구간은 붓스트랩 데이터로 계산된 공정능력지수들을 크기대로 정렬시켜 원 데이터로 계산된 공정능력지수 \mathcal{C}_p 의 위치를 찾아낸 후, 편의가 있으면 이를 조정하여 신뢰구간을 설정하는 방식이다.

4. 모의실험

정규분포, 카이제곱 분포, 스튜던트 t 분

포, 일양분포 등 네 가지 공정분포에 대하여 두 지수 C_{pl} 와 C_{pn} 의 신뢰구간을 SP, PB, ABC 등 세 가지 붓스트랩 방법으로 계산하고, 모의실험을 통해 포함비율로써 포함확률(0.9)을 추정한다. 포함비율의 계산을 위해 각 공정분포에 해당하는 공정능력지수의 참값을 계산하여 C_p^+ 로 나타낸다.

모든 경우에 있어서 공정의 목표치 $T=49$, 관리한계 $USL = 60, LSL = 40$, 손실함수의 모수 $\gamma=2.25, A=8$ 등으로 설정하였다. 표본크기 30과 75 두 가지에 대하여 각각 10,000회씩 실험하여 90% 붓스트랩 신뢰구간과 그 포함비율을 계산하였다. 이 구간에 대한 포함빈도는 모수 $p=0.90$ 와 $N = 10,000$ 을 갖는 이항확률변수이므로 포함확률에 대한 99% 신뢰구간을 계산하면 다음과 같다.

$$0.90 \pm 2.576 \sqrt{\frac{(0.90)(0.10)}{10,000}} \Rightarrow (0.8923, 0.9077) \quad (17)$$

만약 신뢰구간의 포함비율이 위의 구간을 벗어난다면, 그 신뢰구간은 유의수준 1%에서 잘못된 것이라고 말할 수 있다.

각 공정분포에 대한 붓스트랩 모의실험 절차는 다음과 같다.

- (1) 각 분포(정규분포, 카이제곱 분포, 스튜던트 t 분포, 일양분포)에 대한 해당 모수를 정하고 난수를 발생시켜 초기 표본을 구한다.
- (2) 위에서 생성된 표본들을 이용하여, 3가지 형태의 90% 양측 붓스트랩 신뢰구간을 설정한다.
- (3) (2)에서 설정된 신뢰구간 안에 공정능력지수의 참값이 포함되는지를 결정한다.
- (4) 위의 3단계로 이루어진 한 번의 모의실

험 절차를 $N(=10,000)$ 번 반복한다. 이로부터 90% 양측 붓스트랩 신뢰구간 안에 실제 각 공정능력지수의 참값이 포함되는 비율을 계산한다.

4.1 정규분포

공정의 분포가 정규분포일 때의 모의실험 수행결과는 <표 1>, <표 2>와 같다. $\mu = 50, \sigma = 2$ 인 경우(표 1)에는 표본크기 (n)에 상관없이 지수 C_{pl} 의 SB 신뢰구간만 사용가능한 것으로 나타났다. 정규분포임에도 불구하고 지수 C_{pl} 의 신뢰구간의 포함비율은 하한값을 벗어났다.

$\mu = 52, \sigma = 3$ 인 경우(표 2)에는 표본크기 75에서는 지수 C_{pl} 의 SB, ABC 신뢰구간도 사용가능한 것으로 나타났다. 모든 경우에 있어서 지수 C_{pl} 의 신뢰구간이 지수 C_{pm} 의 신뢰구간보다 정확한 것으로 나타났다. 전반적으로 SB 신뢰구간이 가장 정확하고, ABC 방법이 PB 방법보다는 정확한 것으로 나타났다. 전체적으로 지수 C_{pl} 신뢰구간의 폭이 지수 C_{pm} 에 비해 좁게 추정되었으며, 결과적으로 포함비율이 낮게 나타났다. 결론적으로 공정이 정규분포를 따르는 경우에도 표본이 충분히 크지 않은 경우에는 지수 C_{pl} 보다 지수 C_{pm} 의 신뢰구간이 더 정확하며, SB 방법이 가장 우수한 것으로 판단된다. 따라서 비정규분포 뿐 아니라 정규분포를 따르는 공정에도 지수 C_{pm} 을 사용하는 것이 바람직할 것으로 보인다.

4.2 카이제곱 분포

공정의 분포가 카이제곱분포일 때, 모의 실험 수행결과는 <표 3>, <표 4>와 같다. 지수 C_{pm} 에 대한 신뢰구간의 포함비율은 일관성 있게 참값 0.90에 가까운 값을 갖으며, 그 중에서도 SB 방법에 의한 신뢰구간이 가장 정확한 것으로 나타났다. 반면, 자유도가 커짐에 따라 개선되긴 하지만 지수 C_{pl} 는 모든 경우에 있어서 신뢰구간이 좁게 추정되는 것으로 나타났다. 카이제곱 분포와 같이 왜도(skewness)가 큰 분포에 대해서도 지수 C_{pm} 의 붓스트랩 신뢰구간은 바람직한 특성을 갖는 것으로 보인다.

<표 1> 정규분포일 때 90% 붓스트랩 신뢰구간

($\mu = 50 \quad \sigma = 2$)
(포함비율, 신뢰구간의 폭에 대한 표준편차)
신뢰 하한, 신뢰 상한 (10,000회 평균)

	Sample Size (n)			
	30			
	$C_p^+ = 1.7221$			
	C_{pl}		C_{pm}	
$\mu=50$ $\sigma=2$	SB		PB	
	ABC			
75				
SB	PB		ABC	
	ABC			

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같다고 할 수 없음은 의미한다.

<표 2> 정규분포일 때 90% 붓스트랩 신뢰구간

$$(\mu = 52 \quad \sigma = 3)$$

(포함비율, 신뢰구간의 폭에 대한 표준편차)
신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu = 52$ $\sigma = 3$	Sample Size (n)			
	30			
	$C_p^+ = 1.2403$			
	C_{pl}		C_{pm}	
SB	(0.8911*, 0.040)		(0.8950, 0.056)	
	1.145	1.347	1.130	1.378
PB	(0.8767*, 0.038)		(0.8825*, 0.055)	
	1.156	1.354	1.144	1.388
ABC	(0.8850*, 0.037)		(0.8884*, 0.054)	
	1.158	1.357	1.144	1.389
75				
SB	(0.8975, 0.014)		(0.8979, 0.021)	
	1.181	1.304	1.169	1.321
PB	(0.8920*, 0.014)		(0.8932, 0.021)	
	1.184	1.306	1.174	1.325
ABC	(0.8947, 0.014)		(0.8957, 0.021)	
	1.185	1.308	1.174	1.326

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같다고 할 수 없음을 의미한다.

<표 3> 카이제곱분포일 때 90% 붓스트랩 신뢰구간 (자유도 4)

(포함비율, 신뢰구간의 폭에 대한 표준편차)
신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu = 50$ $\sigma = 2$	Sample Size (n)			
	30			
	$C_p^+ = 1.9466$			
자유도=4	C_{pl}		C_{pm}	
	SB	(0.7749*, 0.271)		(0.9020, 0.200)
1.457		2.256	1.627	2.430
PB	(0.8013*, 0.257)		(0.8775*, 0.193)	
	1.553	2.316	1.688	2.471
ABC	(0.7603*, 0.230)		(0.8849*, 0.184)	
	1.534	2.249	1.673	2.439
75				
SB	(0.6107*, 0.112)		(0.9032, 0.071)	
	1.539	2.017	1.741	2.217
PB	(0.6577*, 0.109)		(0.8905*, 0.070)	
	1.574	2.043	1.762	2.233
ABC	(0.6048*, 0.101)		(0.8941, 0.069)	
	1.562	2.014	1.756	2.223

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같다고 할 수 없음을 의미한다.

<표 4> 카이제곱 분포일 때 90% 붓스트랩 신뢰구간 (자유도 10)

(포함비율, 신뢰구간의 폭에 대한 표준편차)
신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu = 50$ $\sigma = 2$	Sample Size (n)			
	30			
	$C_p^+ = 1.8448$			
자유도=5	C_{pl}		C_{pm}	
	SB	(0.8279*, 0.197)		(0.9040, 0.155)
1.480		2.129	1.573	2.235
PB	(0.8346*, 0.188)		(0.8808*, 0.150)	
	1.546	2.172	1.618	2.267
ABC	(0.8125*, 0.170)		(0.8865*, 0.144)	
	1.531	2.127	1.608	2.245
75				
SB	(0.7566*, 0.080)		(0.9003, 0.056)	
	1.561	1.956	1.670	2.067
PB	(0.7843*, 0.079)		(0.8893*, 0.056)	
	1.584	1.973	1.685	2.080
ABC	(0.7493*, 0.074)		(0.8933, 0.055)	
	1.575	1.955	1.681	2.073

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같다고 할 수 없음을 의미한다.

4.3 스튜던트 t 분포

공정분포가 자유도 4, 5, 10 등을 갖는 스튜던트 t 분포인 경우에 대하여 각각 $\mu = 50, \sigma = 2$ 를 갖도록 데이터를 변환하여 실험을 수행하였다. 모의실험 수행결과는 <표 5> ~ <표 7>과 같다. 지수 C_{pm} 은 표본의 크기(n)가 작을 때는 SB 신뢰구간만이 정확성을 보였으나, 표본의 크기가 커지면서 자유도 5 이상에서는 모든 방법에 의한 신뢰구간이 정확한 것으로 나타났다. 반면, 지수 C_{pl} 의 신뢰구간은 자유도와 표본크기 및 방법에 관계없이 모두 부정확한

것으로 나타났다. 다만 자유도가 증가함에 따라 정확성이 개선되었다. 지수 C_{pn} 의 신뢰구간 폭의 표준편차 또한 지수 C_{pl} 신뢰구간 폭의 편차에 비해 전반적으로 작게 나타나 더 안정적임을 알 수 있다.

스튜던트 t 분포는 대칭이며 정규분포와 유사한 형태를 갖기 때문에 큰 표본을 취해 히스토그램을 그려보더라도 육안으로 식별하기 어렵다. 정규분포로 간주하여 지수 C_{pl} 를 사용한다면 부정확한 신뢰구간을 도출할 위험이 있다.

<표 5> 스튜던트 t 분포일 때 90% 붓스트랩 신뢰구간 (자유도 4)
(포함비율, 신뢰구간의 폭에 대한 표준편차) 신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu = 50$ $\sigma = 2$	Sample Size (n)			
	30			
자유도=4	$C_p^+ = 1.8401$			
	C_{pl}		C_{pn}	
SB	(0.8247*, 0.194)		(0.8980, 0.139)	
	1.509	2.129	1.585	2.200
PB	(0.8431*, 0.179)		(0.8794*, 0.135)	
	1.569	2.168	1.624	2.228
ABC	(0.8197*, 0.162)		(0.8833*, 0.131)	
	1.557	2.133	1.615	2.211
75				
SB	(0.8027*, 0.110)		(0.8974, 0.051)	
	1.573	1.973	1.674	2.046
PB	(0.8238*, 0.101)		(0.8874*, 0.051)	
	1.596	1.990	1.688	2.058
ABC	(0.7934*, 0.093)		(0.8910*, 0.050)	
	1.587	1.972	1.684	2.052

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같다고 할 수 없음을 의미한다.

<표 6> 스튜던트 t 분포일 때 90% 붓스트랩 신뢰구간 (자유도 5)
(포함비율, 신뢰구간의 폭에 대한 표준편차) 신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu = 50$ $\sigma = 2$	Sample Size (n)			
	30			
자유도=5	$C_p^+ = 1.8026$			
	C_{pl}		C_{pn}	
SB	(0.8594*, 0.174)		(0.8993, 0.134)	
	1.507	2.083	1.556	2.145
PB	(0.8543*, 0.162)		(0.8773*, 0.130)	
	1.560	2.118	1.594	2.172
ABC	(0.8387*, 0.147)		(0.8843*, 0.125)	
	1.548	2.086	1.585	2.154
75				
SB	(0.8343*, 0.088)		(0.9026, 0.048)	
	1.575	1.941	1.644	2.000
PB	(0.8513*, 0.083)		(0.8943, 0.048)	
	1.595	1.956	1.657	2.011
ABC	(0.8283*, 0.078)		(0.8953, 0.047)	
	1.586	1.940	1.654	2.006

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같다고 할 수 없음을 의미한다.

<표 7> 스튜던트 t 분포일 때 90% 붓스트랩 신뢰구간 (자유도 10)
(포함비율, 신뢰구간의 폭에 대한 표준편차) 신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu = 50$ $\sigma = 2$	Sample Size (n)			
	30			
자유도=10	$C_p^+ = 1.7525$			
	C_{pl}		C_{pn}	
SB	(0.8716*, 0.141)		(0.8982, 0.123)	
	1.512	2.020	1.519	2.067
PB	(0.8607*, 0.135)		(0.8813*, 0.119)	
	1.555	2.049	1.553	2.091
ABC	(0.8587*, 0.126)		(0.8850*, 0.117)	
	1.546	2.027	1.546	2.079
75				
SB	(0.8754*, 0.058)		(0.9049, 0.045)	
	1.584	1.900	1.604	1.937
PB	(0.8781*, 0.057)		(0.8954, 0.045)	
	1.600	1.912	1.616	1.947
ABC	(0.8716*, 0.055)		(0.8964, 0.045)	
	1.594	1.902	1.613	1.943

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같다고 할 수 없음을 의미한다.

4.4 일양(uniform)분포

$\mu = 50, \sigma = 2$ 및 $\mu = 52, \sigma = 3$ 두 가지 경우의 일양분포에 대하여 실험한 결과는 <표 8>, <표 9>와 같다. 이 경우에도 지수 C_{pm} 에 대한 SB 신뢰구간이 모든 경우에서 정확한 것으로 나타났다. 지수 C_{pl} 에 대한 신뢰구간은 모든 경우에서 부정확한 것으로 나타났으며, 표본크기가 클수록 부정확성이 증가하였다. 이는 정규분포에 비해 큰 첨도(kurtosis)를 갖는 일양분포의 특성이 대표본에서 더욱 부각되어 지수 C_{pl} 에 대한 신뢰구간을 왜곡시킴을 의미한다. 이러한 현상은 카이제곱 분포나 자유도가 작은 스튜던트 t 분포에서도 발견할 수 있다.

<표 8> 일양분포일 때 90% 붓스트랩 신뢰구간 ($\mu = 50 \sigma = 2$)

(포함비율, 신뢰구간의 폭에 대한 표준편차) 신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu=50$ $\sigma=2$	Sample Size (n)			
	30			
	$C_b^+ = 1.6562$			
	C_{pl}		C_{pm}	
SB	(0.8798*, 0.089)		(0.8997, 0.099)	
	1.548	1.932	1.465	1.917
PB	(0.8183*, 0.086)		(0.8780*, 0.096)	
	1.573	1.949	1.492	1.936
ABC	(0.8303*, 0.083)		(0.8877*, 0.095)	
	1.570	1.943	1.488	1.930
	75			
SB	(0.7809*, 0.032)		(0.8984, 0.036)	
	1.613	1.843	1.532	1.805
PB	(0.7333*, 0.031)		(0.8904*, 0.036)	
	1.622	1.850	1.541	1.813
ABC	(0.7427*, 0.031)		(0.8942, 0.036)	
	1.621	1.848	1.540	1.811

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같고 할 수 없음음을 의미한다.

<표 9> 일양분포일 때 90% 붓스트랩 신뢰구간 ($\mu = 52 \sigma = 3$)

(포함비율, 신뢰구간의 폭에 대한 표준편차) 신뢰 하한, 신뢰 상한 (10,000회 평균)

$\mu = 52$ $\sigma = 3$	Sample Size (n)			
	30			
	$C_b^+ = 1.2601$			
	C_{pl}		C_{pm}	
SB	(0.8727*, 0.041)		(0.8961, 0.065)	
	1.146	1.350	1.137	1.419
PB	(0.8651*, 0.038)		(0.8841*, 0.064)	
	1.156	1.357	1.154	1.431
ABC	(0.8703*, 0.038)		(0.8868*, 0.063)	
	1.157	1.358	1.154	1.431
	75			
SB	(0.8398*, 0.013)		(0.9044, 0.024)	
	1.182	1.304	1.181	1.352
PB	(0.8445*, 0.012)		(0.8981, 0.023)	
	1.185	1.306	1.187	1.357
ABC	(0.8482*, 0.012)		(0.9023, 0.023)	
	1.186	1.307	1.187	1.357

* 유의수준 1%에서 신뢰구간의 포함확률이 참값 0.9와는 같고 할 수 없음음을 의미한다.

5. 결론

본 연구에서는 SB, PB, ABC 등 세 가지 붓스트랩 신뢰구간 추정 방법을 사용하여 EDF에 기초한 공정능력지수 C_{pm} 의 신뢰구간을 추정하고, 모의실험을 통하여 포함확률을 추정하였다. 정규분포 가정에 기초한 지수 C_{pl} 의 신뢰구간과 비교한 결과, 지수 C_{pm} 의 SB 신뢰구간은 공정의 분포 유형 및 표본크기에 영향을 받지 않고 정확한 포함확률을 갖는 것으로 나타났다. 특히, 공정이 정규분포를 따르는 경우에도 소표본에서는 지수 C_{pm} 의 신뢰구간이 C_{pl} 의 신뢰구간보다 더 정확한 것으로 나타났다.

세 가지 붓스트랩 방법 중 SB 방법이 모든 공정의 분포에 대해 수행결과가 우수하였다. 따라서 공정이 정규분포를 따른다고 확신할 수 없는 현장에서 지수 C_{pm} 과 SB 신뢰구간을 사용하면, 비교적 정확한 공정능력 평가가 가능할 것으로 판단된다.

추후 연구로는 비대칭 손실함수에 기초한 공정능력지수와 그 신뢰구간에 대한 분석이 기대된다.

참고문헌

- [1] 임태진, 송현석(2003), “EDF 기대손실에 기초한 로버스트 공정능력지수”, 「품질경영학회지」, 31권, 1호, pp.109-122.
- [2] 정영배, 문혜진(2002), “역정규 손실함수를 이용한 공정능력지수에 관한 연구”, 「품질경영학회지」, 30권, 3호, pp.66-78.
- [3] 조중계, 김주성, 박병선(1999), “공정능력지수 C_{pk} 에 대한 보다 나은 비모수적 붓스트랩 신뢰구간에 관한 연구”, 「응용통계연구」, 12권, 1호, pp.45-65.
- [4] Boyles, R. A.(1991), “The Taguchi Capability Index”, *Journal of Quality Technology*, Vol. 23, No. 1, pp.17-26.
- [5] Chan, L. K., Cheng, S. W., and Spring, F. A.(1988), “A New Measure of Process Capability : C_{pm} ”, *Journal of the Quality Technology*, Vol. 20, No.3, pp.162-175.
- [6] Chan, L. K., Xiong, Z., and Zang, D.(1990), “On The Asymptotic Distributions of Some Process Capability Indices”, *Communications in Statistics- Theory and Methods*, Vol. 19, pp.11-18.
- [7] Choi, K. C., Nam, K. H., and Park, D. H.(1996), “Estimation of Capability Index Based on Bootstrap Method”, *Microelect. & Reliab.*, Vol. 36, No.9, pp.1141-1153.
- [8] Franklin, L. A. and Wasserman, G. S.(1992), “Bootstrap Lower Confidence Limits For Capability Indices”, *Journal of Quality Technology*, Vol. 24, No 4, pp.196-210.
- [9] Efron, B. (1979), “Bootstrap Methods: Another Look at the Jackknife”, *Ann. Statist.*, Vol. 7, No. 1, pp.1-26.
- [10] Efron, B. (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York London
- [11] Hall, P. (1988), “Theoretical Comparison of Bootstrap Confidence Intervals”, *Ann. Statist.*, Vol. 16, pp.927-953.
- [12] Pearn, W. L., Kotz, S., and Johnson, N. L.(1992), “Distributional and Inferential Properties of Process Capability Indices”, *Journal of Quality Technology*, Vol. 24, No. 4, pp.216-231.
- [13] Spiring, F. A.(1993), “The Reflected Normal Loss Function”, *The Canadian Journal of Statistics*, Vol. 21, No. 3, pp.321-330.
- [14] Spiring, F. A. and Yeung, A. S.(1998), “A General Class of Loss Functions with Industrial Applications”, *Journal of Quality Technology*, Vol. 30, No. 2, pp.152-162.